present in the query sequence, as well as a modified ("masked") version of the query

sequence in which all the annotated repeats have been masked (*e.g.*, replaced by Ns).

The RepeatMasker program is publicly available (*see, e.g.*, the website at

repeatmasker.genome.washington.edu/).

Please amend the paragraph bridging pages 7-8 of the application as

follows:

Other usable programs include Censor (Jurka, *et al.* (1996) *Computers and*

Chemistry 20:119-122; *see, e.g.*, the website at girinst.org/Censor_Server.html; Genetic

Information Research Institute, California); Satellites or Repeats (Institut Pasteur, Paris;

see, e.g., the website at bioweb.pasteur.fr/seqanal/interfaces); and others.

Please amend the last full paragraph on page 9 of the application as

follows:

Typically, the masked sequence (*i.e.*, collection of selected subsequences)

will be compared with the genome database using a suitable algorithm such as BLAST

(*see, e.g.*, the BLAST server at the National Center for Biotechnology Information). A

BLAST or equivalent search will identify sequences within the genome that are

homologous to the masked sequence, preferably ranked in order of similarity to each

subsequence.

Please amend the paragraph bridging pages 10-11 of the application as

follows:

A preferred example of algorithm that is suitable for determining percent

sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms,

which are described in Altschul *et al., Nuc. Acids Res.* 25:3389-3402 (1977) and Altschul

et al., J. Mol. Biol. 215:403-410 (1990), respectively. BLAST and BLAST 2.0 are used,

with the parameters described herein, to determine percent sequence identity for the

nucleic acids and proteins of the invention. Software for performing BLAST analyses is

publicly available through the National Center for Biotechnology Information. This

algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying

short words of length W in the query sequence, which either match or satisfy some

positive-valued threshold score T when aligned with a word of the same length in a

database sequence. T is referred to as the neighborhood word score threshold (Altschul

et al., supra). These initial neighborhood word hits act as seeds for initiating searches to

find longer HSPs containing them. The word hits are extended in both directions along

each sequence for as far as the cumulative alignment score can be increased. Cumulative

scores are calculated using, for nucleotide sequences, the parameters M (reward score for

a pair of matching residues; always > 0) and N (penalty score for mismatching residues;

always < 0). For amino acid sequences, a scoring matrix is used to calculate the

cumulative score. Extension of the word hits in each direction are halted when: the

cumulative alignment score falls off by the quantity X from its maximum achieved value;

the cumulative score goes to zero or below, due to the accumulation of one or more

negative-scoring residue alignments; or the end of either sequence is reached. The

BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the

alignment. The BLASTN program (for nucleotide sequences) uses as defaults a

wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both

strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength

of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff &

Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation

(E) of 10, M=5, N=-4, and a comparison of both strands.

Please amend the paragraph bridging pages 12-13 of the application as

follows:

Typically, the primers will be designed not only based on the size of the

product, but also taking into account any of a large number of considerations for optimal

primer design, *e.g.*, to exclude potential secondary structures within the primers, with a

desired T_m (that is preferably similar for each member of a pair of primers), to include

additional sequences such as restriction sites to facilitate cloning of the amplified

product, etc. Examples of suitable programs for designing (and analyzing potential

primer sequences) include, but are not limited to, Primer3 (from the Whitehead Institute;

website at genome.wi.mit.edu/cgi-bin/primer/primer3.cgi, PrimerDesign website at

chemie.uni-marburg.de/~becker/pdhome.html Primer Express® Oligo Design Software

(PE Biosystems), DOPE2 (Design of Oligonucleotide Primers website at

dope.interactiva.de/); DoPrimer (website at doprimer.interactiva.de); NetPrimer (website

at premierbiosoft.com/netprimer.html); Oligos-U-Like--Primers3 (website at

path.cam.ac.uk/cgi-bin/primer3.cgi); Oligo (v5.0); CpG Ware™ Primer Design Software,

PrimerCheck (website at chemie.uni-

marburg.de/~becker/freeware/freeware.html#primercheck), and others. General

parameters for designing primers can be found in any of a large number of resources and

publications, including Dieffenbach, et al., in PCR Primer, A Laboratory Manual,

Dieffenbach et al., Ed., Cold Spring Harbor Laboratory Press, New York (1995), pp.133-

155; Innis, et al., in PCR protocols, A Guide to Methods and Applications, Innis, et al.,

Ed., CRC Press, London (1994), pp. 5-11; Sharrocks, in PCR Technology, Current

Innovations, Griffin, H.G., and Griffin, A.M, Ed., CRC Press, London (1994) 5-11.

IN THE CLAIMS:

Please cancel claim 39.

Please amend claims 11-12 and 27-28 and add new claims 40-42 as

follows:

11.   (Once Amended) The method of claim 1, wherein said first process

is executed using [Repeat Masker software] a software program that screens sequences

for:

i.   interspersed repeats that are known to exist in mammalian

genomes and;

ii.   low complexity DNA sequences.